Section A

Q1. Almost all the class got part of this incorrect, part correct. The gradient is calculated anew for a new loss function - the one stated has gradient 2(f-y), not just (f-y). Most people however correctly stated the nested loop for SGD. Several people forgot to state the update for parameter t.

Q2. Most of the class (>50%) got this correct - a few (about 20%) could name it but not state it correctly in mathematics.

Q3. Several people (about 30%) did not get this correct as they mis-interpreted what the probabilities were used for. If the distribution of X was {0.2,0.2,0.1,0.5}, the values of the X are not 0.2, 0.1, etc - these are the *probabilities* of the values in a discrete distribution.

Section B

Most students performed well, or very well, on Ross King's question.